

## **CATEGORISING VOCALISATION OF ENGLISH /l/ USING EPG, EMA AND ULTRASOUND**

A.A. Wrench & James M. Scobbie

Speech and Language Sciences,  
Queen Margaret University College, Edinburgh

**ABSTRACT:** We consider allophonic and speaker-specific variation in the alveolar gestures found in vocalised and consonantal /l/. EMA or ultrasound give tongue position data and EPG gives simultaneous alveolar contact data. We find systematic vocalisation with interspeaker variation in the prosodic distribution of these vocalised forms. The data suggest some allophonic variation is subtle, continuous and gradient, while some is more categorical.

### **INTRODUCTION**

Many varieties of English display a readily perceptible (and sometimes strikingly large) allophonic variation in /l/ largely determined by its syllabification. Such “clear” vs. “dark” patterns have been examined articulatorily primarily in standard American English /l/ (Giles & Moll 1975; Sproat & Fujimura 1993; Narayanan et al. 1997). Syllable onsets condition the clear variant, and codas the dark.

These studies show English /l/ can have more than one lingual constriction. The influential study by Sproat & Fujimura (1993), moreover, proposed that both the relative strength of the different articulations and their relative intergestural timing were determined by the position of the /l/ in the syllable, and suggested this to be the cause of the clear/dark allophony.

The “stronger”, more occlusive of the constrictions of /l/ is due to a midsagittal apical or laminal gesture towards an anterior target, usually the alveolar ridge (or upper teeth). Even when there is alveolar contact, the free flow of air through the oral cavity means that /l/ is classified as an approximant. To take a simple intervocalic context, for example, the tongue tip is extended upwards and forwards while the edges of the tongue blade are compressed inwards, bringing them away from the upper teeth. The midsagittal section of the tongue may dip in a “saddle” shape behind the tip. The resulting complex vocal tract airway that typically branches over both sides of the tongue into a substantial sublingual space. These resonating cavities generate the typical acoustic correlates of a lateral.

While alveolar contact seems to be cross-dialectally typical of English /l/ in an onset position, the approximant nature of /l/ is far more apparent in coda positions. Here, the “weaker” non-occlusive dorsal articulation of /l/ towards the velar, uvular or pharyngeal regions (Narayanan et al. 1997), seems to be relatively more prominent. Indeed, along with an enhancement of the dorsal gesture which narrows the posterior airway in dark /l/, the “strong” alveolar gesture may be weakened so radically — in some dialects at least — that there is no contact between active and passive articulators. In such a case the /l/ is said to be “vocalised”. In such dialects (cf. Foulkes & Docherty 1999), vocalisation is typically said to be accompanied by strong labialisation so that /l/ in these contexts is like [u] or [o]. Moreover, most descriptions imply vocalisation is categorical. Such transcriptions reflect the deletion or extreme reduction of the alveolar gesture, enhanced pharyngeal constriction and labialisation.

Given that in studies of clear/dark /l/ the consonantal gesture of alveolar contact is articulatorily stronger in onset position and weaker in coda position, and given that Giles & Moll (1975) observed some vocalisation in fast speech and adjacent to low vowels, we hypothesise that dark /l/ and vocalised /l/ may be closely related as coda forms. We would not wish to claim that they are endpoints, nor on a continuum, but the a priori similarity of the two forms merits close empirical investigation.

Sproat and Fujimura’s analysis suggests rather a different emphasis: namely that /l/ allophony is not consonant-specific, but rather is due to general processes of gestural weakening in the coda. These processes diminish the more “consonantal” gestural components of multigestural segments (e.g. /m/, /w/, /l/), and delay them relative to the more “vocalic” gestures which may be boosted in strength. This position has been defended in previous work which has looked for and found similarities between /l/ allophony and syllable-based allophony in other consonants (e.g. Browman & Goldstein 1995).

Proceedings of the 6th International Seminar on Speech Production, Sydney, December 7 to 10, 2003.

In this paper, we report new findings from an on-going project looking at interspeaker variation in the systems of /l/ allophony visible from examination of the alveolar/anterior gestures recorded during production of /l/. In Scobbie & Wrench (2003) we reported our findings on word-final /l/. We found that all eight of our speakers vocalised /l/ in pre-labial codas, and that there was systematic /l/ sandhi. That analysis was reliant on a binary categorisation of EPG data into consonantal vs. vocalised tokens. In this paper, we report on a larger range of contexts, and add results on gradient tongue tip height from simultaneously-collected EMA data, as well as some pilot results on tongue-shape using ultrasound.

## METHOD

The data comes from the MOCHA (Multi-Channel Articulatory) database, in which each speaker reads 460 phonetically rich sentences, providing data on more than 10,000 phones per speaker (Wrench 2000). In addition to acoustics, simultaneous articulatory channels comprising Carstens EMA, Laryngograph and EPG are recorded. The "Reading" EPG palate provides tongue-palate contact data at 62 normalised positions. The sampling rate of EPG in the MOCHA corpus is 200Hz. Previously-released data for SW, SS and AS, unchecked data for the others, the list of sentences, and MATLAB analysis tools can be found online at <http://bell.qmuc.ac.uk>.

## Subjects

The subjects analysed here are chosen because their data happens to be available, not on any linguistic grounds. They have varied accents. Five are English: SS, SW, near-RP speakers from SE England; AS, a SE England speaker with impressionistically obvious rounded vocalised /l/; JW, an RP speaker from the north of England; and SK with a slightly more pronounced northern accent. One speaker, FS, is typical of west central Scotland (e.g. Glasgow). Finally, JN and AT are from the USA, and do not have a particular regional accent.

## Materials

We examine a total of 11 specified contexts, each exemplified by between 12-20 examples from the sentences, giving 172 tokens of /l/ per speaker (less 7 unanalysable, so  $n=1365$ ). The EMA analysis is of 4 of these contexts, giving 120 tokens per subject (of which 6 were unanalysable, so  $n=474$ ).

- Three contexts are preconsonantal. Actually, they are all prelabial in order that we can interpret any lingual contact as being due to the /l/ and not, for example, the /k/ in "milk" or the /d/ in "hold". In one word-final case /l/ is "strong", i.e. following a tonic (lexically stressed) vowel (e.g. "peel"). In the other word-final case /l/ is "weak" i.e. following a lexically unstressed vowel (e.g. "apple"). The third case is word-internal prelabial /l/, and all these cases are strong (e.g. "elm"). These are all coda contexts.
- Two contexts (weak and strong) are prepausal. These word-final contexts are also codas.
- In six contexts the /l/ is prevocalic. In one case the /l/ is word-initial, clearly an onset. In two (strong and weak) it is word-final before another word which itself starts with a vowel, which means the /l/ may resyllabify postlexically to onset. In three cases the /l/ is word-medial: either preceding a fully stressed vowel (e.g. "alone"), an onset, or preceding a weak vowel, in which case it may be weak (e.g. "bungalow") or strong (e.g. "yellow"). The former context is of doubtful syllabic status, whereas the latter is usually said to be ambisyllabic.

The other aspects of the lexical and phonological context are not controlled, resulting in token-token variation in parts of the context not mentioned. The dorsal coil data therefore is too noisy to be useful, due to the varied contextual vowel quality. Consequently we look only at the more stable tongue tip gesture. The materials are available at <http://sls.qmuc.ac.uk/dept/jscobbie/home.htm>.

## Analysis

The MOCHA analysis is based on an examination of EPG contact patterns in an alveolar zone on the palate using MATLAB analysis tools. This zone of 18 contacts covers the front three rows, excluding the 1<sup>st</sup> and 8<sup>th</sup> edge columns. If any contact occurs in the zone, the token is classified as consonantal.

MATLAB tools also find the tangential velocity minimum of the EMA tongue "tip" coil (about 7-10mm posterior to the tip) associated with an /l/ constriction (cf. the circle and dashed line in Figure 1a). This point defines the target of /l/, at which the tongue tip position (TT x & y) is recorded. Comparable TT data can thus be obtained both from consonantal /l/ tokens (seen in Figure 1a) and vocalised /l/ tokens

Proceedings of the 6th International Seminar on Speech Production, Sydney, December 7 to 10, 2003.

which lack EPG contact, or in cases with no clear TTy peak (not shown). Note that in Figure 1a, the left (“L”, red) and right (“R”, green) vertical cursors mark the first and last frames of EPG contact.

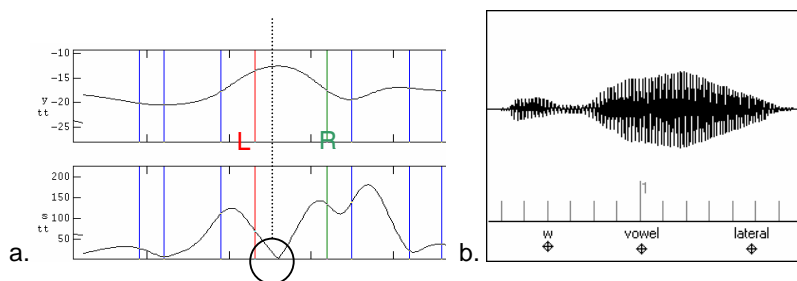


Figure 1. Left upper: TT coil vertical “y” location (mm). Left lower: the speed (mm/s) of the TT coil. Horizontal tick marks are at 100ms intervals. Right: “a wall”, marked by 50ms ticks. See text for more details.

Figure 1b shows the annotation points for the ultrasound example presented below in Figure 6. The simultaneous ultrasound video pictures (24Hz) and EPG (100Hz) are digitised and analysed in a version of Articulate Assistant modified to allow annotation and analysis of video images (e.g. ultrasound), audio, and EPG. The annotation of [wɔɪ] reflects max., min. and max. labialisation.

### MOCHA EPG-EMA RESULTS

Figure 2 presents mean vocalisation rates. Pooling all speakers, onsets are typically consonantal: 96% of word initial and 98% of word medial cases have contacts in the alveolar zone. Word-final /l/ which is prevocalic is also typically consonantal, except for subject JN (see Scobbie & Wrench 2003 for more details). Pooling the other seven subjects, only 8% are vocalised in the strong context and 5% in the weak context. In codas, however, things are quite different. Pre-labial coda /l/ is *vocalised* in 94% of strong word final cases, 93% of weak word final cases and 95% word medially. Vocalisation is not simply coda weakening, however. Half of the subjects do not vocalise strong pre-pausal /l/, while their pre-labial /l/ is vocalised, yet both of these contexts are prosodic codas. Note also that prosodic context and metrical strength interact for two speakers, so that pre-pausal /l/ is vocalised more when it is weak (syllabic) than when strong. Finally, word-medial intervocalic /l/ are typically consonantal (apart from JN's), so that for the other seven speakers, ambisyllabic /l/ in strong position (eg “yellow”) is only 11% vocalised and in weak position (eg “bungalow”) the mean is 5%.

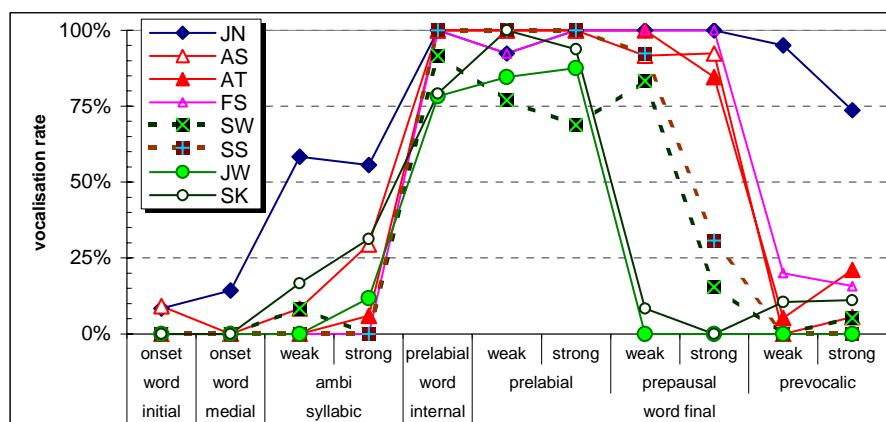


Figure 2. Vocalisation rates, based on existence of EPG contact.

Apart from the case of JN, the differences between most contexts for the other speakers appear generally categorical. Perhaps four sandhi systems can therefore be identified, each of which involves increasing amounts of vocalisation. JN vocalises all positions bar consonantal word-initial and word-medial onsets. FS, AS and AT vocalise all codas. SS and SW vocalise prelabial codas and weak prepausal ones. JW and SK vocalise only prelabial codas, and not prepausal ones.

Different prosodic contexts condition different gestural targets for different speakers, creating a range of systems. Further, for four speakers, the variation between different types of coda seems as marked as the onset-coda variation itself. The range of speakers' systematic behaviour suggests that a simple set of general principles governing the phonetic interpretation of syllabic affiliation is not a major factor in the specification of the fine level of detail which each speaker's gestural planning seems to require.

Important though alveolar contact is for [l], interpreting the results is difficult because they may unduly reflect the binary categorisation method. Therefore, intrinsically gradient data from EMA is presented to give more information. Figure 3 shows speakers' mean tongue tip height ("TTy"), based on all tokens whatever their EPG classification, normalised on a speaker-specific basis. As well as adding new information about the extent of tongue lowering, these results underline the systematic presence and absence of alveolar contact revealed above for different speakers and environments.

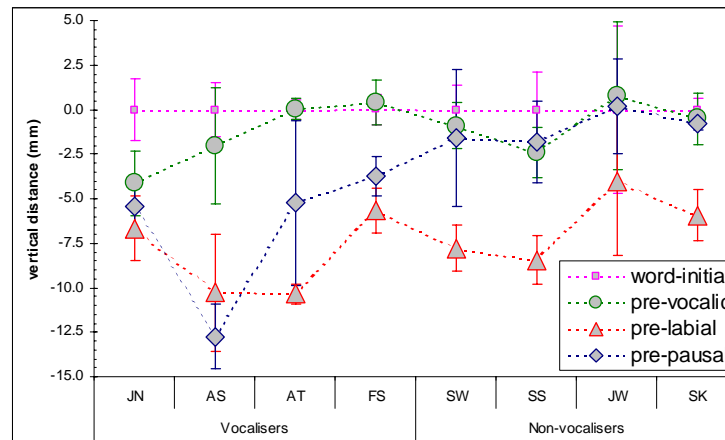


Figure 3. Mean vertical height of TT coil of word-final strong /l/ in prevocalic, prelabbial and prepausal contexts (n=19, n=16, n=13) relative to word-initial /l/ (=0mm) (n=12). Whiskers indicate one standard deviation.

Figure 3 also shows correspondences in TTy in different syllable positions. The mean TTy location of the "non-vocalisers" SW, SS, JW and SK is congruent in prepausal (coda) and word-final prevocalic (postlexical onset/ambisyllabic) position. A high rate of vocalisation need not be due to extreme tongue lowering, however, as can be seen from JN's results. Note that AS's variants are the most extreme, with TTy more than 10mm lowered in both coda contexts. This may be systematic, and reflect the fact that he, uniquely among the subjects, has a perceptibly vocalised (and labialised) /l/.

#### ULTRASOUND-EPG RESULTS

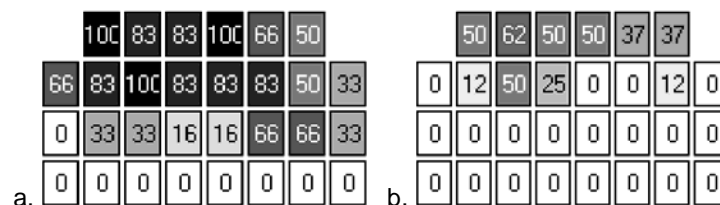


Figure 4. Average maximum tongue-palate EPG contact patterns in four anterior rows during /l/ from (a) "real afters", and (b) "real puddings". Each cell indicates the % of the tokens in which there was a contact in that cell.

JW's EMA data above suggest underlying gradient vocalisation, while her EPG data reveal a complex and rather categorical system. In some pilot experiments combining EPG and ultrasound we look again at her productions. We collected six tokens of word-final /l/ in a prevocalic context and eight in prelabbial context. JW knew that we were interested in her /l/ productions, and perhaps this self-awareness explains why she vocalised only 3 of 8 tokens, rather than the 88% value seen in Figure 2.

First, the EPG results (Figure 4) show that the prevocalic context conditions a greater area of contact, as well as more frequent and consistent contact. The preconsonantal /l/ in Figure 4b reflects weaker contact patterns for consonantal /l/ plus the effect of the three tokens with no contact at all. Second, ultrasound provides good evidence of tongue tip raising and a saddle in the prelabial case (Figure 5b), even in the vocalised tokens. In the prelabial case (Figure 5a) four tokens have a clear saddle shape. We presume the sampling rate of 24Hz and the relatively short contact duration (mean 58ms, s.d. 13ms) contribute to a failure to capture the alveolar contact of the relatively rapid onset-like /l/ in two tokens. The mean contact duration of prelabial /l/ was longer at 78ms (s.d. 24ms) for the five consonantal cases, giving more opportunity for ultrasound to capture the lateralised tongue shape.

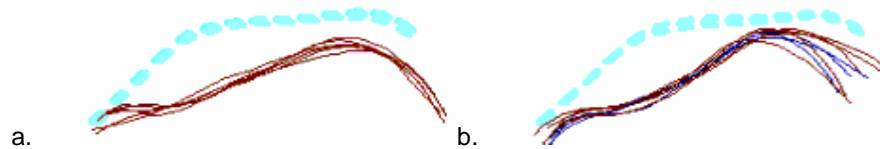


Figure 5. Ultrasound data on tongue shape during /l/ from (a) “real afters”, (b) “real puddings”, showing superimposed splines of the midsagittal tongue surface (solid lines) at the time of EPG maximum constriction for [l] (red), or during the vocalised /l/ (blue). The palate (dashed line) spline was estimated from frames in the data showing JW swallowing. Anterior to left.

We infer from these results that JW does indeed have subtle gradient weakening. For such speakers some contexts are likely to result in so much gestural undershoot that vocalisation occurs, while other factors such as prepausal lengthening (only for strong syllables in the case of SW and SS) may counteract this tendency. We do not wish to imply, however, that her behaviour is *exclusively* due to general patterns of coda weakening and prepausal lengthening and the like, for we have no reason to doubt her system and the fine phonetic detail of her targets is anything other than dialect-specific.

For balance, we also exemplify a speaker with much more radical vocalisation in prepausal position, namely the second author JS (from west central Scotland). It seems he has a system like FS or AS above, with quite extreme vocalisation and apparently categorical sandhi. Unlike AS, neither JS nor FS have strong labialisation in vocalised /l/ (i.e. it is [ɫ], not [u]). Ultrasound & video data of the lips were collected to let us examine the extent of vocalisation, and also to check that there is no labialisation.

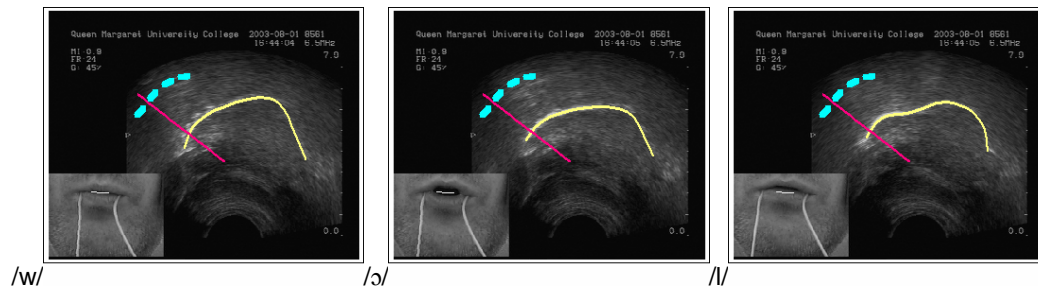


Figure 6. Ultrasound of tongue and video of lips in prepausal “wall”. Anterior is to left. Hard palate is in dashes, tongue in yellow. Reference bars for tip position (pink) and [w] labialisation (white) are discussed in the text.

JS’s vocalisation is indeed extreme: while JW’s vocalised tokens tended to be further from the alveolar ridge (circa 3mm vs. 2mm), JS’s tongue is on average 24mm from the alveolar ridge (n=4) on the reference dimension indicated in pink in Figure 6. The tip is as low for JS’s /l/ as it is for his [w], accompanied by pharyngealisation rather than [w]’s velarisation. Without showing any clear signs of tongue tip raising, /l/ still has a saddle in the tongue. This may be due to a “lateral” shaping of the tongue blade independent of any actual alveolar constriction, and/or due to the pharyngealisation. Figure 6 also exemplifies the fact that JS’s dark vocalised /l/ tokens do have labialisation. See the white bar which indicates the width and position of labial opening in the [w]. JS’s vocalised /l/ does not involve protrusion and rounding as seen in his [w]: rather, the lower lip alone is raised.

## CONCLUSION

Giles & Moll (1975) found that dark /l/ may be articulatorily vocalised in American English, typically in faster speech rates and following low vowels. Narayanan et al. (1997) found very little vocalisation in their examination of dark /l/, perhaps because the use of MRI required sustained [l] sounds. The sentential data from the MOCHA corpus provides relatively unmonitored data, and our speakers, with a variety of standard accents, vocalise frequently and systematically, suggesting vocalisation is a natural aspect of the clear/dark dichotomy more along the lines of Giles & Moll's results. Moreover, we add to previous findings by revealing complex prosodic conditioning of alveolar contact of singleton /l/, giving rise to various subtle and extreme systems of allophony and external sandhi.

A pilot ultrasound investigation of heavily vocalised /l/ revealed a lingual saddle despite no evidence of tip raising. These more extreme types of vocalisation (e.g. JS, AS) may therefore not be suited for analysis within a single model of gestural undershoot, but demand a more categorical approach. For JW and the other speakers with less extreme vocalisation, on the other hand, we need to account for the relatively categorical patterns of EPG contact being due to underlying gradient weakening.

While we concur that there are articulatory tendencies such as undershoot in weak contexts, speakers of different accents of English can acquire, by definition, different fine phonetic details of allophones in linguistically-identical environments. Such subtle variation within a given context means the role of the context itself must not be exaggerated as a determining factor. Dark and vocalised /l/ are, we think, related to each other, and to general properties of the coda. Whether there is a *continuum* from dark to vocalised /l/ is a far more complex and subtle problem which requires much more data on interspeaker variation, especially from vernacular varieties of English which are seldom represented experimentally. While Giles & Moll (1975) had access to (and made very good use of) x-ray cinematography, more recent research has had to make do with often inferior non-invasive techniques. We hope that ultrasound (in combination with acoustics, EPG and video) will offer a non-invasive and safe source of articulatory data to take these issues forward.

## ACKNOWLEDGEMENTS

The MOCHA corpus was funded under EPSRC Grant GR/L78680 to the first author. The second author is funded under ESRC Fellowship R000271195. Financial support for ultrasound was from the Scottish Higher Education Funding Council (SHEFC).

## REFERENCES

- Browman, C.P. & Goldstein, L. (1995) "Gestural Syllable Position Effects in American English" In *Producing Speech: Contemporary Issues. For Katherine Safford Harris, F. Bell-Berti and L. J. Raphael* (Eds.), 19-34. New York, NY: AIP Press.
- Foulkes, P. & Docherty, G. (eds.) (1999) "Urban Voices: Accent Studies in the British Isles". London: Edward Arnold.
- Giles, S.B. & Moll, K.L. (1975) "Cinefluorographic study of selected allophones of English /l/" *Phonetica* **31**, 206-227.
- Narayanan, S.S., Alwan A.A. & Haker, K. (1997) "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals" *Journal of the Acoustical Society of America* **101** (2), 1064-1077.
- Scobbie, J.M. & Wrench, A.A. (2003) An articulatory investigation of word final /l/ and /l/-sandhi in three dialects of English" *Proceedings of the 15th International Congress of Phonetic Sciences*, 1871-74.
- Sproat, R. & Fujimura, O. (1993) "Allophonic variation in English /l/ and its implications for phonetic implementation" *Journal of Phonetics* **21**, 291-311.
- Wrench, A.A. (2000) "A multi-channel/multi-speaker articulatory database for continuous speech recognition research" *Phonus* **5**, 1-13.